DOCUMENT RESUME

ED 202 455 IR 009 301

AUTHOR	Steinberg, Esther R.
TITLE	Using PLATO To Teach Problem Solving.
INSTITUTION	Illinois Univ., Urbana. Computer-Based Education Research Lab.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, Md.
REPORT NO	CERL-R-E-18
PUB DATE	Aug 80
CONTRACT	NICHHD-HD-0591
NOTE	65p.
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Computer Assisted Instruction; *Decision Making; *Educational Strategies; Feedback; *Instructional Innovation; Primary Education; *Problem Solving; Teacher Effectiveness
IDENTIFIERS	*PLATO IV

ABSTRACT

This report describes the results of three studies which investigated the use of PLATO to teach problem solving. Subjects were 244 children from kindergarten through third grade. The first study explored the extent to which kindergartners and second graders used the knowledge about which answers were right and which were wrong to develop the correct strategy to solve a novel picture problem. The next study looked at the procedures used by kindergartners and second graders while they were in the process of trying to get the answer. The third study considered three conditions of teaching first and third graders the necessary problem solving skills: experience only, feedback consisting of a visual record of the student's work on the first five problems, and interactive instructional feedback on how to solve each problem that the child had done wrong. Abstracts, methods, results, and discussions are given for each of the studies. Results of this research led to a post hoc study which is described. Seven figures show sample displays and nine references are also given. (CHC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED202455

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Using PLATO® to Teach Problem Solving

Esther R. Steinberg

Computer-based Education Research Laboratory
University of Illinois at Urbana-Champaign

August 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Bitzer

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR009301

Copyright © August 1980
by the Board of Trustees, University of Illinois

All rights reserved. No part of this book may be reproduced in any form or by any means without permission in writing from the author.

Portions of this work were supported by Contract No. HD-0591 of the National Institute of Child Health and Human Development and a grant from the UIUC Research Board.

Acknowledgments

I wish to thank the children, teachers, and principals of Kenwood, Columbia, and Westview elementary schools in Champaign for their cooperation.

The help of Julie Garrard and Heather Greenberg in preparing these manuscripts is also much appreciated.

TABLE OF CONTENTS

I. Executive summary 1

A. Purpose and Rationale 1
B. Procedure 1
C. Results 2
D. Implications for teaching problem solving on PLATO 3

II. Study 1: Evaluation processes in young children's problem solving . . 7
Abstract 8
Preliminary study 11
Experiment 20

III. Study 2: Children's judgments of sufficient information 25
Abstract 26
Experiment 1 28
Experiment 2 35

IV. Study 3: Feedback Compared to Practice Alone as Instruction in Problem Solving 37
Abstract 38
Experiment 40
Revised Feedback and a Post Hoc Study 50
problem solving.

V. Combined references 52

VI. Figures 54

Using PLATO to Teach Problem Solving

EXECUTIVE SUMMARY

A. Purpose and Rationale

Some problems, such as column addition, can be solved by following an algorithm, that is a specific set of rules. Others, such as finding a defective component in an electronic system, are not routine and require more general reasoning and problem solving skills. Students of all ages, particularly novices in a subject matter area, encounter great difficulty when faced with the latter types of problems. Although instructors give explanations and show examples before the students try to do the problems, students frequently continue to do them wrong on their first attempts. They may not fully understand the preliminary instruction because they lack familiarity with the problem. They may understand only the specifics of the examples, but not be able to infer the general strategy or its implications for related problems. They need some sort of feedback to help them develop appropriate problem solving strategies. The PLATO author cannot assume that all students will gain insights merely from experience with the problem, or a game-like version of it. Therefore, the PLATO author must think about what kinds of help to provide. Should the author simply tell whether the answer is right or wrong? why it is wrong? tell the student the right answer? explain why it is right? The purpose of this research was to get data that are needed to answer these questions.

In this series of studies I have made two fundamental assumptions that are based on current thinking in psychology. (1) Immature problem solvers do not necessarily follow the same problem solving processes as the experts. In order to provide students with meaningful assistance, it is necessary to study the processes they use when first attempting to solve the problem (Resnick, 1976; Case, 1978; Gagne & White, 1978). (2) Problem solving and reasoning skills may not be entirely age related, but also very much dependent on cognitive maturity and problem solving experience (Brown & Deloache, 1978). Therefore, an understanding of how children develop some of these skills should enable us to develop guidelines for teaching students who are older, but who are inexperienced in a particular subject matter area.

B. Procedure

Overview. These studies evolved from my observations that kindergartners were unable to solve a problem in letter discrimination, which was one of their reading tasks on PLATO (Obertino, P., Fillman, L., Gilfillan, J., Silver, D., & Yeager, R., 1977). The purpose of the first study was to find out why this was so. Was it due to the children's inability to generate a strategy? Were they simply getting answers by trial and error, or was there evidence of systematic use of some stategies? Did they shift to different strategies if the ones they were using resulted in wrong answers? Thus, the goals of the first study were to investigate (1) children's ability to generate stategies and (2) their ability to evaluate their

strategies after they generated an answer and found out it was wrong. The second study focused on their ability to make inferences while they were solving the problem rather than after they derived and answer as in the first study. The purpose was to find out if the children were able to judge before completing the problem, whether the information they had obtained was sufficient to solve the problem. The results of these first two studies led to some hypotheses about feedback that should be effective for teaching the correct strategies. A test of these hypotheses was the purpose of the third study. Results of this research led to a post hoc study, in which effective instruction resulted from a combination of initial instruction and feedback which enabled the student to see the functional relationship between the strategy he had used and the results of that strategy.

To summarize, two studies characterized students' problem solving procedures. One considered students' procedures after they had obtained a wrong answer; the other, while they were in the process of trying the get the answer. The third study considered different ways of teaching them the necessary problem solving skills.

Subjects. Participants were 244 boys and girls, from kindergarten through third grade, in Champaign public schools.

Materials. The research medium was PLATO. The basic task was as follows. Two pictures were displayed simultaneously. They were alike in all respects except one. For example, one picture showed a camel with its head up, the other with head down (Figure 1, see p. 54). Simultaneously a third covered picture, was displayed and purported to be identical to one of the other two. The problem was to be solved in two steps: uncover just half of the covered picture, and then decide which picture it would match for sure. Obviously, the correct strategy was to uncover the part that was not the same, in order to obtain discriminative information.

In another version, the task employed two-digit numbers such as 21 and 28. In its most complex form, the problem involved three three-digit numbers, such as 145, 175, and 138 (Figure 2, see p. 55).

C. Results

In the first study, solving the problem was defined as doing eight problems in a row correctly. Students as young as five-year olds did generate strategies but only a few solved the problem. Some even shifted to different strategies when the ones they were using proved unsuccessful. But they were unable to generate an appropriate strategy. It turned out that this was due to the task structure, which allowed them to get the right answer, by chance, 70% of the time. They were right so much of the time and wrong so little, there was not much reason to evaluate their strategies. PLATO's special capabilities were used to change the task so that students could not get the right answer when they used the wrong strategy. Using this structure, two-thirds of the children were successful.

The second study showed that kindergartners did not judge whether they had sufficient information while they were trying to solve the problem. Even when they had the option of saying that they did not have enough information to obtain the correct answer, they did not use it. Second graders seemed able to judge when they had insufficient information, but sometimes went on to complete the problem anyhow. They knew that they couldn't solve the problem but they just wanted "to see what would happen."

The following were some of the students' difficulties. They misunderstood the problem. They generated strategies that worked only in special cases. They could not predict the consequences of their strategies; that is, they could not anticipate the answer they would get as a consequence of their procedure.

The third study compared practice above to different kinds of feedback. Feedback that led the students through the correct steps for each incorrect problem, was no better than just practice doing the problems. Even after PLATO showed them how to do the problem, they went back to their own incorrect strategy. They apparently did not understand the implications of what PLATO was telling them to do. Just going through the steps, putting in correct answers did not result in their making inferences about the underlying strategy.

Feedback that was a visual record of the first five problems, the student's answers, and whether they were right or not, was no better than just practice doing the problems. The students were unable to make inferences about why the wrong answers were wrong. "No" meant "I did it wrong." Period. Nothing more.

For the less complex (two-digit or pictures) tasks, successful students generated a strategy that was specific to that task. They had to generate a different, more general strategy when faced with the more complex task, three-digit numbers.

In the third study, successful instruction was developed. Success was attributed to several factors. First, the teaching problem was not at the simplest level. Instruction was given for the three-digit, rather than the two-digit task. The students were thus prevented from generating a strategy that was generally incorrect, but just happened to result in the correct answer for 2-digit problems. Second, the students were given initial instruction to be sure they understood the problem. They were taught how to predict all of the numbers that could possibly be in each part of the covered number. Finally, those alternatives were kept in full view while the student tried to solve the problem (Figure 3, see p. 56). Thus, the student could see on a single display what all the choices were as well as the results he got from his choice.

D. Implications for teaching problem solving on PLATO.

Although the research reported here was limited to a single type of problem and subject population (5- to 8-year olds), the results nevertheless point to critical issues and questions that the PLATO author

should consider when writing problem solving lessons for other subject matter areas and diverse subject populations.

1. The structure of the task

   a. Be sure the students understand the problem. You might check on this by asking some leading questions. You might let them do a problem, and ask questions at various intermediate steps. Another alternative is to let them try to solve a problem and then ask questions to determine whether their understanding of the problem is the same as that which you intended.

   b. Check to see if there is some other element in the task that allows the student to get the correct answer by lucky guess. If so, consider using PLATO to generate the correct answer contingent on the student's procedure. See the experiment in Study 1 for an example of how this can be done.

2. Feedback

   A primary goal of feedback is to help the student understand the problem solving procedure or strategy, not just to help him or her get the answer to a particular problem. The trick is to figure out how to do so. You want to give the student a chance to do as much of the problem as he can on his own; yet you do not want to leave him "stuck" and unable to procede. You don't want the feedback to be such obvious information that the student already knows it; yet you don't want it to be so advanced that the student does not understand the implications for solving the problem at hand. The point is that feedback ought to be available at the level of the students' understanding.

   One way to implement this is to give feedback, as needed in increasing amounts of detail. For example, the first time your feedback could be a general hint, like "Before you try to do the problem again, formulate a plan or strategy." If the student does the problem wrong a second time, suggest, "Think of alternative ways you can try to solve this problem. Try to predict the likelihood of success of each of them." In even more detailed feedback you could suggest several procedures and have the student predict the one most likely to succeed. Notice that I am suggesting that at first your feedback might be statements of general problem solving techniques. If the student needs more hints, go on to problem-specific suggestions.

   Feedback can also be provided while the student is trying to solve the problem. At some intermediate step, you might ask the student, or help him consider whether the information he has obtained is sufficient to solve the problem, or whether he should start all over before trying to complete it.

   A form of feedback that proved useful in my research is displaying on the screen several alternative ways of doing the problem, and at the same time the results of the student's choice of how to

do it. It enables the student to literally see the consequences of his actions while at the same time keeping in mind the other actions he might have taken. It is a matter of setting the conditions so that the student can figure out the essential features of the problem, or be able to discriminate between the essential and non-essential aspects.

Evaluation Processes in Young Children's Problem Solving

Esther R. Steinberg

University of Illinois
at Urbana-Champaign

This work was supported, in part, by Contract No. HD-0591 of the National Institute of Child Health and Human Development. The cooperation of students and teachers in participating schools of Champaign, Illinois is gratefully acknowledged.

Abstract

This study investigated the extent to which kindergartners and second graders used the knowledge about which answers were right and which were wrong to develop the correct strategy to solve a novel picture problem. The second graders evaluated their answers completely; that is they generated the correct strategy and solved the problem. Only 2 of the 23 kindergartners in the preliminary study did so. Seventeen of them shifted strategies in response to wrong answers, but did not generate appropriate ones; four children perseverated in a single, wrong strategy, even in the face of incorrect answers. In the main experiment 23 out of 35 kindergartners generated the correct strategy. The difference in the proportion of successful problem solvers was ascribed to the fact that in the main experiment, the child could never get the answer right if the strategy was wrong.

Evaluation Processes in Young Children's Problem Solving

This investigation characterizes the processes by which children try to develop a correct strategy during repeated attempts to solve a problem. It evolved from the author's observations that kindergartners were unable to do a problem in letter discrimination which was one of their reading tasks on PLATO, the University of Illinois system for computer-assisted instruction. The author might have concluded that the children did not solve the problem correctly because it was simply too difficult for kindergartners, that they lacked the skills and strategies that are needed to solve it. However, she took the position that children this young are not as inept as we may think (Gelman, 1978), and that a modified version of this task could serve as a vehicle for studying the children's strategic behavior.

This work, like the original problem, was done on PLATO. Some of its unique characteristics were used to manipulate the task in ways that are impossible, or at best difficult, via traditional media. Therefore in order to give the reader a better understanding of this research, first it is necessary to give a brief description of the features of the PLATO system that are relevant to this study.

A PLATO terminal includes an 8" (.20 m) square display panel and a typewriter-like keyset. Information is displayed on the panel in orange, against a black background. Actually, the display panel consists of a matrix of 512x512 dots. The computer shows characters or pictures on the panel by lighting up the appropriate dots within 8x16-dot rectangular areas. The student communicates by typing responses on the keyset. The terminal can also be equipped with a touch sensing device so that the student can respond by touching the panel with a finger. The lesson author is not limited to using the standard letters that are provided by the system, but may design his or her own characters such as a bird or a rectangle of lighted dots.

In the reading problem under consideration here, the student communicates with the computer by touching the screen with a finger. A "hidden" letter on the screen is identical to one of two letters displayed at the top of the screen (Figure 4 , see p. 57). The student uncovers the hidden letter piece by piece until he or she thinks that there is enough information to match it to one of the letters at the top of the screen (Obertino et al., 1977). It should be noted that on PLATO, the stimulus that is uncovered is generated after the subject makes a response. What appears to be a covered letter is simply a rectangular block of lighted orange dots. The letter per se is not actually physically present. After the subject signifies which part of the screen to uncover by touching that part, the computer turns off the rectangle of lighted dots and then displays the appropriate part of the letter. It will be shown below how this feature of PLATO was used in the main experiment.

Notice the skills and the strategies that are necessary to solve the problem. The mature problem solver must be able to make inferences on the basis of partial information. Other skills needed are: (1) the ability to

predict which information will be useful for making such an inference, and which will not; (2) the ability to subdivide the whole into parts, and then reconstruct the whole; and (3) the ability to mentally subdivide each letter into 6 pieces, or areas, in the same way as the hidden letter and locate a particular relative area to uncover.

Kindergartners are deficient in these skills and strategies, according to both Piagetians and other psychologists. Piaget (1976) contends that children this young are incapable of the kind of deductive reasoning that is required here. Furthermore, the problem requires a reversible kind of thinking in which the problem solver not only mentally divides the whole letter into a subset of 6 pieces, but also recreates the whole from these pieces. This is the kind of thinking that is also supposedly beyond the 5-year old (Inhelder & Piaget, 1964). Some psychologists note that such young children have encoding or attentional deficiencies (Gholson & McConville, 1974; Siegler, 1976). Thus in this task kindergartners may fail to attend to the critical features, to the similarities and differences between the letters. It would seem, then, that the task should be discarded as one which is too difficult for kindergartners.

The task, however, was not discarded. First, it is not the case that we can generalize about what children can and cannot do on the basis of Piagetian type tasks. Brainerd (1978) has pointed out that there is little or no direct evidence that it is beneficial, or even appropriate, to apply the results of cognitive-developmental research to instructional practices. Second, success in problem solving is sensitive to task structure and several features of the reading task may have contributed to the children's failure in it. In the reading task, there were six parts to scan and this may have been too many to keep track of for children this young. Furthermore, the children were not forced to try to solve the problem by uncovering only 1 or 2 areas. They were allowed to uncover as many pieces as they wished, so they simply transformed the problem into a familiar, matching task, by uncovering the entire letter and then matching to one of the given letters. Finally, letters are abstract characters for kindergartners and it is not easy for them to find a label to identify and remember the part of the letter they wish to look at.

The task was modified to overcome these difficulties. It was anticipated that even in the revised problem, initially the students would choose unsuitable strategies and make errors, but it was hoped that they would ultimately solve it. The purpose of this research was to characterize the students' evaluation of their answers and their strategies in the process of trying to obtain the correct strategy.

The following modifications were made. Pictures of familiar objects were used instead of letters. This enabled the child to label part of it. For example, the child could think about and refer to the tail or the head of an animal when deciding which part of the hidden figure to uncover. The pictures were divided into 2 parts instead of 6, for less scanning. Also, the student was allowed to uncover one and only one part

of the covered picture. In contrast to the original problem, the one used in this research forced the student to make a match decision on the basis of incomplete information.

To summarize the actual problem then, two pictures were displayed simultaneously on the PLATO screen. They were alike in all respects except one. For example, one picture showed a camel with its head up, the other with head down (Figure 1, see p. 54). Simultaneously a third, covered picture, was displayed and purported to be identical to one of the other two. The problem was to be solved in two steps: uncover or peek at just half of the covered picture, and then decide which picture it would match for sure ("peek" and "match" responses, respectively). Obviously, the right strategy was to obtain discriminative information by peeking at the part that was not the same.

How will a young child try to solve such a problem? Let us assume that the student discerns the critical information, that the pictures are alike in one half and not in the other. Let us also suppose that the child generates a typical 5- year old plan, position preference, and peeks at the left side of every picture . Since not all of the pictures are different in the left half, the child will see the nondiscriminative part (e.g. the tail) of some of the pictures, and will not be able to tell for sure which view will match the entire covered picture. Therefore it will be necessary to guess, and thus the match will be wrong some of the time. The question is, to what extent will the child evaluate these wrong answers? There may be no evaluation at all. Even when the match is wrong, the child might maintain the same peek strategy on every item (Level 1). Another possibility is that after one or more incorrect matches, the response to these wrong answers will be to abandon the current peek strategy in favor of some other, inappropriate plan. For example, a shift might be made from a position preference to alternating sides, or to a win-stay, lose-shift system. This would indicate an understanding of the need to make some revisions, but inability or failure to generate a suitable strategy (Level 2). Finally, the child might evaluate completely. In response to the fact that some of the matches are wrong, a better understanding of the problem might be sought. The child would then use this information to find a more suitable alternative, one that would ultimately result in solving the problem correctly (Level 3).

Preliminary Study

Method

Subjects. The kindergartners were all 24 children in a morning section of a public school in a middle class socio-economic area. Data of one child were incomplete and hence deleted. Data were reported for 13 girls and 10 boys, ages ranging from 5-0 to 5-8 with mean 5 years, 5 months. It was expected that although the problem would be difficult for five-year-olds, it could be solved readily by older children. In order to verify this assumption, a class of second graders was included in the study. They attended a school in the same system but in a lower

income area. Again the data for one child were incomplete and therefore omitted, leaving 19 children, 12 boys and 7 girls, ages ranging from 6-11 to 8-7 with mean 7 years, 6 months. Three of the children were one year below grade level for their age.

All children had previous experience with PLATO, since two terminals were a regular part of each classroom.

Materials. The experimental medium was PLATO. Twelve pairs of pictures were generated by computer line drawings. Six of the sets were top-bottom (three top, three bottom) different, and the other six were different left-right (three of each). The pictures were: camel, teddy bear, dragon, cup, vase, rabbit, bird, indian, lady, girl, boy. Each picture was framed by a 2 1/8 in. (.05 m) square. A horizontal (or vertical) line was drawn across the middle to emphasize the type of difference (Figure 1, see p. 54).

Two display arrangements were used. In the horizontal display, the two views of the picture were shown side by side with 1 in. (.025 m) between them. A covered (all orange) square was displayed midway between them and 1 in. (.025 m) below. A horizontal (or vertical) line divided it into two parts. In the vertical display, the two pictures were shown one above the other and the covered picture was midway between them to the right.

The computer randomly generated the order of the pictures. The computer also randomly chose which view would be displayed in each position and which would be the covered one. For example, the computer randomly chose the camel with the head up to be displayed on the left, the camel with the head down on the right, and the camel with the head down as the hidden one. In order to prevent superstitious behavior from developing, the randomness was subject to the constraint that neither the correct match nor the discriminative peek would be in the same position on more than three successive trials.

Design. Each subject participated in two sessions. The first session consisted of an orientation and 12 trials. The second session consisted of 12 trials, using the same pictures as in the first session, but in a different, random presentation order. Thus the total number of trials was 24 for each child. Since the choice of the hidden view was randomly generated, the correct match for a picture set was not necessarily the same in both sessions. For example, the camel with the head down could be the hidden one at one session but the camel with the head up could be correct the next session.

Procedure. Each child was randomly chosen by the experimenter, invited to come to a PLATO terminal that was in the classroom, and to sit down beside the experimenter. The first session began with the orientation. The initial display showed two rabbits which differed only in the right side of the picture. The experimenter asked the child if the rabbits were exactly alike. If the child responded correctly, the experimenter asked how they were different. If the child answered

incorrectly, the experimenter told the child to notice the ears and paws, and asked the question again. Every child answered correctly after this prompt. The next display (kangaroos) was like the problems in the trials and included the covered picture. The student was asked if the kangaroos were exactly alike, and after the child agreed they were not, the experimenter explained that a kangaroo exactly like one of them was hiding under the "cover." The child was told that the task was to identify which one it looked like by peeking at just half of the hidden picture. The student touched the part he or she wished to see; the computer erased that part and showed the appropriate part of the kangaroo. If the strategy was incorrect, the experimenter commented that the uncovered part of the kangaroo looked like both kangaroos and the child could only guess which one it matched. The student decided which picture the covered one would match, and then touched it. The computer revealed the rest of the covered kangaroo. The student was asked if the answer was right, that is, if the pictures matched.

The student then began Trial 1 and followed the same procedure for each of the 12 trials. For each trial, the two views of a picture and the covered picture were displayed. The student chose one part to peek at and touched it; the computer uncovered that part; the child touched the picture it would match, or look like "for sure"; the computer uncovered the rest of the partially covered picture; the student checked to see if the answer was right. Notice that once the picture was fully displayed, it was the student, not the experimenter, who judged the accuracy of the match. The entire task was self-paced. All six pictures of one kind were shown before the other set was begun. For example, all 6 pictures which were left-right different, like the camel, were shown. Then the set of pictures which were top-bottom different were presented. Each such set of 6 was called a phase. The order of items within a phase was randomly generated. The initial display arrangement and type of picture difference was counterbalanced across students. After the student completed all 12 trials, a message appeared on the screen and was read by the experimenter, "That's all for today. Thank you (student's name)."

The second session was given on the next school day. This session was identical to the first, except that the student saw the other display arrangement and the trials were differently ordered and presented, as noted above.

Results

A trial was considered correct if and only if both the peek and matth responses were correct. Criterion for success was 10 out of 12, or 5 successive trials correct in a session. Only 1 of the 23 kindergartners reached criterion after the first session and another did so after the second. Twelve of the 19 second-graders reached criterion in the first session and 6 others in the second session. Every child who reached criterion in the first session also reached criterion in the second.

A student was said to maintain a strategy in a phase if he used the same strategy on at least 5 of the 6 trials in that phase. (Recall that all items in a phase were either left-right, or top-bottom, different.) If the student's strategy in a phase could not be discerned, it was labeled undefined. Operational definition of levels of evaluation were then generated in terms of strategy maintenance and quality as follows. Level 1: the student maintained the same inappropriate strategy in all four phases. Level 2: the student maintained a strategy in at least one phase, shifted at least once to another strategy or to an undefined strategy in another phase, but did not ultimately solve the problem. Level 3: the student solved the problem in at least one session. Four kindergartners were at Level 1, 17 at Level 2, and 2 at Level 3 (Table 1). The numbers were 1, 0, and 18, respectively, for second graders. The one child at Level 1 was repeating second grade.

Table 1

Level of Evaluation After Two Sessions

Group	n	Level 1 None	Level 2 Partial	Level 3 Complete
Preliminary Study				
Kindergarten	23	4	17	2
Second Grade	19	1	0	18
Experiment				
Kindergarten	35	4	8	23

Kindergartners at Level 2 employed the following strategies: position preference on the peek, such as top; position preference on the match; win-stay, lose-shift; alternate peeks, such as peek left, peek right; and peek at the part that is the same in both views, exactly the opposite of the correct strategy. Since there were 4 phases per child, for all 17 children at Level 2 there were a total of 68 phases. For these childen the strategies listed above were observed 22, 6, 9, 5, and 2 times, respectively. The strategies were undefined in the other 24 phases.

Mean number of correct items, out of a possible 24, was 11.30 for kindergartners, and 18.58 for second graders (Table 2). This difference was significant, $t(40) = 7.85$, $p < .01$.

Table 2

Mean Number of Correct Answers After Two Sessions

Group	n	Mean	Standard deviation
Preliminary study			
Kindergarten	23	11.30	2.53
Second Grade	19	18.58	3.47
Experiment			
Kindergarten	35	15.17	4.48

The child could match correctly by chance, whether the peek strategy was right or wrong (peeked at the discriminative part or not). Thus, the final response could be right, even though it was just luck and the child really did not know what he or she was doing. In fact every kindergartner and all but one second-grader made at least one such lucky guess. The mean number of lucky guesses was 5.9 for kindergartners and 2.8 for second-graders.

Thirteen of the 23 kindergartners and 5 of the 19 second-graders mismatched; that is, they peeked at the side that provided discriminative information, but did not correctly identify the matching pictures. The mean number of mismatches was 1.51 (s.d.=1.26) for kindergartners and .26 (s.d.=.45) for second graders. The difference between the means was significant, $t(40) = 4.426$, $p < .01$. Overall, after a correct peek, kindergartners matched with 90% accuracy and second graders with 99%. To put it in terms of mismatches, the mean proportion of correct peeks followed by mismatches was .10 (s.d.=.13) for kindergartners and .01 (s.d.=.03) for second graders. This difference was also significant, $t(40) = 3.23$, $p < .01$.

Latency for match response, that is time between first (peek) and second response was recorded for each trial. Means were calculated, by grade and accuracy of response (Table 3). Analysis of variance revealed that both main effects, grade and response value, were significant, $F(1,78) = 5.10$, $p < .05$ and $F(1,78) = 43.90$, $p < .01$, respectively. Interaction was also significant, $F(1,78) = 4.09$, $p < .05$. Children in both grades took more time to match after an incorrect peek than after a correct one. However, the time difference was 2.7 seconds for the kindergartners compared to 5.1 seconds for second graders.

Table 3

Mean Time (in Seconds) Between Peek and Match Responses for Kindergartners and Second Graders

Group	Response Value of Peek					
	Correct			Incorrect		
	n	Mean	Standard deviation	n	Mean	Standard deviation
Kindergarten	23	8.22	1.53	23	10.92	3.32
Second grade	18	5.71	.53	18	10.78	3.78

Analysis of Variance

Source	df	MS	F
Group (G)	1	35.418	5.10*
Response Value (RV)	1	305.105	43.90**
G x RV	1	28.428	4.09*
Within Cell	78	6.951	

*p<.05
**p<.01

Some of the children spontaneously verbalized their strategy. Six kindergartners and one second grader generated the following strategy. First guess which picture it is; then peek to verify. If the display reveals discriminative information, match the correct picture even if it means altering the initial guess; if the part revealed is not discriminating, stay with the initial guess because it does partially match, ignoring the fact that the other picture matches that part too. The second grader and one of the six kindergartners ultimately shifted away from this strategy and solved the problem.

At the second session, two of the second graders and one kindergartner remembered some of the correct matches from the first session and assumed they would again be correct at the second session. For instance, when the stimulus was the bunny, the kindergartner said, "It was that one yesterday," pointing to the picture that was recalled. Thus, memory was employed to solve the problem. The child drew on experience from traditional instruction, and assumed that the displays would stay the same, as in a textbook. The child had no way of knowing about the random generating capability of the computer. Unfortunately, in both cases the match was wrong, much to the consternation of the child. The experimenter explained the situation, and none of the children employed the memory strategy on succeeding trials.

The second graders reacted to each wrong match with such spontaneous comments as "oh shoot", "oh no", and "ooh, I should have looked at the other part." Kindergartners, on the other hand were willing to tolerate more than a single wrong answer before doing something about it. The experimenter asked one kindergartner what a person should do about a wrong answer, and the child replied, "If you don't win you just don't care."

Discussion

The problem was difficult for the kindergartners. Just two solved it. At the other extreme, four others simply did not make any inferences. These children apparently evaluated only the final answers (I was right or wrong), and not their strategy. Right or wrong, they peeked in the same position throughout. The rest of the children understood the need to change strategies, and did so at one time or another, but did not choose appropriate ones. After some number of wrong answers among the rights, they evaluated the strategy as unsatisfactory, so they chose another one. However, it is not clear why they did not find the correct one. It may not have been part of their repertoire. On the other hand, it may be that kindergartners do not try to match the strategy to the problem. If one strategy doesn't work, they just try another until they find the right one or get tired or bored and quit searching for alternatives.

Six children used partial information to verify an initial guess whether this was logically justifiable or not.

All but one second-grader solved the problem. Seven of the 18 successful students must have preplanned the strategy because they were

successful from the first item. But the other 11 did not preplan and yet succeeded.

Both kindergartners and second graders apparently did realize that they could not make a match on the basis of an incorrect peek, since they took more time to match after an incorrect peek than after a correct one. However, the fact that the difference between match times after correct and incorrect peeks was so much less for kindergartners than for the second graders (2.7 vs. 5.1 seconds, respectively) suggests that the children in the two grades reacted differently to the incorrect peeks. Since second graders eventually solved the problem, one can assume that, after peeking, they made the inference that the peek was not useful for making a match. Kindergartners, on the other hand, apparently did not infer that the peek was incorrect, and assumed they could match correctly anyhow. As a matter of fact, the task conditions were such that they could match correctly by chance even though the peek was wrong.

This chance aspect of the task may have confounded the results. The number of correct matches the child saw was greater than the number counted correct by the experimenter. The experimenter counted an item correct if and only if both the peek and match responses were correct. But the child could match correctly, by luck, even if the peek was incorrect. Thus, even if the child was using an inappropriate strategy or just guessing, the probability of getting the match response right was .70. That is the case because if the child happened to peek right, $p=.5$, the probability of the match being right was $p=.9$. (Recall that overall match accuracy after a correct peek was 90%.) If the peek happened to be wrong, $p=.5$, the probability of matching right was $p=.5$. Altogether, the probability of a correct match after a correct peek OR an incorrect peek was $.5 \times .9 + .5 \times .5 = .70$. It seems reasonable that this high proportion of chance correct matches gave the kindergartners confidence in their strategies. Although they did switch strategies, it was certainly not after a single wrong match response. Second graders, on the other hand, overcame incorrect strategies even with the 70% probability of getting the correct answer. The difference between the grades may be that these kindergartners did not have the correct strategy in their repertoire. It may also be that kindergartners are more willing than second graders to accept occasional wrong answers. That is, if an answer is wrong a second grader immediately tries to figure out what to do about it. A kindergartner, on the other hand, is willing to accept the fact that not every answer will be correct and is not so quick to seek alternative strategies.

About half (13) of the kindergartners obtained discriminative information from their peek but matched it to the incorrect picture. However, the mean number of mismatches was only 1.5. Therefore, it is not clear whether they lacked the prerequisite discrimination skills that are necessary to solve this problem, or were simply a bit careless.

To summarize, the kindergartners did not solve the problem. It was hypothesized that this was due to either a deficiency in prerequisite skills or to the task conditions, which enabled the students to get the right match by chance even though the strategy was wrong.

Therefore a definitive experiment was conducted to determine whether kindergartners could solve this problem if the probability of getting a correct answer by chance were decreased.

The following revisions were made:

1. A 10-problem discrimination pretest was developed on-line. It was added to test the assumption that kindergartners do have the discrimination skills prerequisite to the task and that their mismatches were not due to a deficiency here. Each problem in the pretest displayed three pictures, two alike and one different. The task was to touch the picture that was not the same. Five of the items were pictures. The other items were one each of shape and color discrimination, two of numbers (15, 51 and 23, 13) and one letters (b, d).

2. In the orientation in the preliminary study, the kangaroo problem was essentially a warmup trial, but it also included feedback from the experimenter. This may have been inadequate for a complete understanding of the task. The student did not have the opportunity of testing out and seeing the results of each of the alternatives (peek at the tail and peek at the head). Therefore, the orientation was changed. The student successively uncovered each part of the kangaroo, and answered questions. That is, attention was drawn to the concept of looking at corresponding parts of the pictures and observing the results.

3. The probability of a lucky correct answer was decreased to .45. In the preliminary study, the computer randomly chose the view that would be the covered one before the child began the problem. However, in this experiment the computer was programmed to choose the covered view after the child finished responding. For example, suppose the child chose to see the tail (nondiscriminative part) in the camel problem. Then if the child decided it would match the camel with the head down, the camel under the cover turned out to have its head up. If the strategy was wrong, the student's final response was wrong.

To understand how this was possible, it is necessary to recall some special characteristics of PLATO. The computer shows characters or pictures by lighting up the appropriate dots within 8x16-dot rectangular areas. It is possible to light up all the dots in an area, giving the psychological effect of covering up a character when in fact the character no longer exists on the screen as a separate entity. Suppose, for example, the author displays the letter, e, on the display and then "covers" the e. What happens is that all of the dots in the rectangle light up, not just the ones that trace out the e. Then the e as such does not exist. The student probably assumes the e is there and is just covered up because that is the experience with traditional media. If a letter or word is covered in a book, the letter does not cease to exist just because it is out of sight. But in the experiment, when part of the picture is purported to be "covered", there is really no picture under it. When the student indicates which part to uncover, the computer first turns off the lighted dots in that area, and then generates the picture by lighting up the appropriate

ones. Since the computer does not generate the picture that is supposedly under the cover until after the student's response, the computer can also wait to make the decision about which view to show until after the student's response. The essential difference between the preliminary study and this experiment is that in the former, the computer decided on the covered view before the student's responses; in the latter that decision was made after, and contingent on the student's responses.

Experiment

Method

Subjects. The children were all of the 37 kindergartners from both morning and afternoon classes of one teacher in another middle-class school in the same school system. Data were deleted for one child who did not complete the experiment and one who was partially blind. There were then 17 boys and 18 girls, ages 5-1 to 6-4, mean age 5 years, 9 months. As in the first experiment, PLATO terminals were a regular part of the classroom and the children had had previous experience with them.

Procedure. The first session began with the pretest, followed by the revised orientation and 12 trials, as in the previous experiment. The second session was given on the next school day and was identical to that in the previous experiment.

Results

Thirty-one children scored 70% or better, on the discrimination pretest, and 22 of them later reached criterion on the problem itself. Four children scored below 70% on the pretest, and only one of them later succeeded on the main task.

On the task itself, 23 of the 35 children reached criterion. This was a significantly greater number, χ^2 (1) = 16.25, $p < .01$, than kindergartners in the preliminary study. Of these 23, 6 succeeded during the first 6-trial phase and 4 in the second phase of the first session. The other 13 children succeeded during the second session the next day.

Four kindergartners were at Level 1, 8 were at Level 2, and 23 were at Level 3 (Table 1). The proportion of children at each level was significantly different from both kindergartners in the preliminary study χ^2 (2) = 16.77, $p < .001$, and from second graders, χ^2 (2) = 6.8, $p < .05$.

All children at Level 1 showed position preference. Two of them had scored below 60% on the discrimination pretest.

The strategies employed by the 8 students at Level 2 were analyzed. Position preference was again most common, with 12 uses, followed by peeking at the side that looked the same (6), win-stay,lose-shift (2), alternate peek position (2), alternate match (1), correct strategy (1), and undefined (8).

Mean number of items correct was 15.17 (Table 2). This was significantly greater than for kindergartners in the preliminary study, $t(56) = 4.189$, $p < .01$, but significantly less than for second graders, $t(52) = 2.88$, $p < .01$.

Mean time between first and second response was computed for all children. Both successful and unsuccessful students spent more time studying their peek response when it showed non-discriminative information than when it showed discriminative. Analysis of variance (Table 4) showed significant main effects as well as a significant interaction between success status and correctness of response $F(1,66) = 9.81$, $p < .01$. Both students who reached criterion and those who did not, averaged about six seconds after a correct response; but after incorrect responses successful students averaged 14.33 seconds compared to 9.71 seconds for unsuccessful students.

These latency differences for successful and unsuccessful students are parallel to the results for second graders and kindergartners, respectively, in the preliminary study (Figure 4, see p. 57). Since latencies in that study included all students, successful and unsuccessful, they were not strictly comparable. Therefore, for purposes of comparison, they were recalculated for each grade, omitting the one unsuccessful second grader and the two successful kindergartners.

Discussion

Only 8 of all 58 kindergartners in both investigations did not evaluate their results at all, but perseverated in a single strategy or behavior throughout (Level 1). The other 51 children were at Levels 2 or 3; that is, most did shift their problem solving strategy when faced with the fact that their final answers were sometimes wrong. However, not all of the kindergartners generated an appropriate strategy (Level 2). The studies yielded significantly different proportions of children who ultimately solved the problem (Level 3). In the experiment, in which the children could not get the final answer right unless they used an appropriate strategy, two-thirds succeeded. In the preliminary study, in which the children could get the final answer right by chance, the proportion was .09.

Although the vast majority (89%) of the kindergartners scored 70% or better on the discrimination pretest, only 71% of that group later solved the problem. This suggests that kindergartners do have the prerequisite discrimination skills for this task, but that such skills are not a sufficient condition for solving the problem. Therefore, for most of the children in the preliminary study, failure to solve the problem cannot be ascribed to discrimination skill deficiencies.

Furthermore, this difference in performance in the two investigations cannot be attributed either to the priming effect of the discrimination pretest or to the change in orientation in the experiment. Only 6 of the 23 successful children reached criterion within the first, 6-trial phase which came immediately after the pretest. Four succeeded only after the second phase and thirteen others did not succeed until the second session

Table 4

Mean Time (in Seconds) Between Peek and Match Responses for Successful and Unsuccessful Students

Group	Response Value of Peek					
	Correct			Incorrect		
	n	Mean	Standard deviation	n	Mean	Standard deviation
Successful	23	6.08	.51	23	14.37	5.73
Unsuccessful	12	6.32	.68	12	9.25	1.69

Analysis of Variance

Source	df	MS	F
Group (G)	1	93.841	8.10*
Response Value (RV)	1	496.086	42.83*
G x RV	1	113.639	9.81*
Within Cell	66	11.582	

*$p<.01$

the next day. If the pretest or revised orientation served as a prompt, one would have expected all 23 children to succeed in the first phase, or at least the first session. Therefore, the difference in results must be attributed to the difference in probability of getting a match right by luck.

A partial explanation for the differences is that kindergartners are not likely to evaluate a strategy if the match answer is right, but the may do so if it is wrong or if a number of answers are wrong. The kindergartners got 70% of the answers right in the preliminary study even if they did not know what they were doing. They did very little strategy evaluation because they got so few problems wrong. In the experiment, however, only half of the answers could be right by chance. More problems were wrong so there were more instances in which strategy evaluation could take place.

In addition to the difference in the number of wrong answers, there was also a difference in the consistency of the feedback the child was getting, and this difference can account for facilitation of problem solving in the experiment. Suppose that initially most of the children tried to solve the problem by hypothesis testing, in typical kindergarten fashion (Mosher & Hornsby, 1966). They guessed which view the covered picture would match, then peeked to verify. If they happened to peek at the nondiscriminative part, the succession of events that followed was different in the two studies. In the first, an incorrect peek could be followed by a correct match or an incorrect match with equal probability, but in the second, an incorrect peek was always followed by an incorrect match. Thus, the experiment enabled the child to make the observation that the match was always wrong when he or she had peeked at the part that was "the same." This consistent pairing of peeking at the part that looked the same occuring with the wrong match drew attention to the peek as the critical element for solving the problem. Most of the children made this observation and subsequently were able to make the inference to peek at the discriminative part.

Kindergartners may indeed be hypothesis testers initially, but they can shift to a more appropriate strategy if the task condition fosters the shift. In some tasks, where there is a finite number of answers to pick from, like the 20 questions type of task, they just keep trying answers until they get the right one, or get tired of the game. In such a task the response may be wrong, but the system of hypothesis testing is not wrong, just inefficient. There is no reason for the child to abandon this strategy other than efficiency, and a child this young is not concerned with efficiency. In the present task, not only was the response wrong, but the strategy of hypothesis testing was also wrong. Hypothesis testing could not ultimately lead to success. In order to succeed the children had to find a different strategy. The children in the preliminary study were unable to come up with the appropriate strategy, but the consistent feedback in the experiment enabled children to do so.

A point that requires further explication is why the children who did switch strategies did not ultimately generate a successful one. One

explanation is that although they did evaluate the peek strategy, it was done as a separate activity and was unrelated to the child's overall strategy. They tried to figure out how to peek without making an adequate inference about how the peek was related to getting the correct match. Another possibility is that after obtaining nondiscriminative information, they were aware that they could not tell with certainty which picture it matched, but did not know what to do about it. It may also be that they thought they could tell with certainty, even though they could not. The time data from both investigations show that unsuccessful students took significantly less time than successful ones to pick a match after non-discriminative information had been obtained from a wrong peek. Apparently successful students knew that they could not make such a match with certainty, and hesitated to respond. However, one cannot conclude that unsuccessful students failed to notice that something was awry, because they did take more time after an incorrect peek than after a correct one. More information is needed to determine the extent of their knowledge about the value of the information they had obtained.

Children's Judgments of Sufficient Information During Problem Solving

Esther R. Steinberg

University of Illinois
at Urbana-Champaign

This work was supported, in part, by a grant from the UIUC Research Board. The author gratefully acknowledges the cooperation of the principals, teachers, and children in the Champaign Public schools who participated in this research. The assistance of Charles Secolsky is also appreciated.

Abstract

Thirty-six kindergartners and forty-two second graders tried to solve a novel problem either with or without the option, during problem solving, of judging whether they had obtained sufficient information to do so. A significantly greater proportion of second graders solved the problems and did so more quickly than kindergartners. Kindergartners did not judge sufficiency of information during problem solving, even with the presence of a prompt. Successful second graders did judge sufficiency during problem solving.

Children's Judgments of Sufficient Information During Problem Solving

By the time children are 5 years old, they are likely to use some kind of systematic plans or strategies to solve a problem (Siegler, 1978). With development, these strategies become increasingly sophisticated (Brown & DeLoache, 1978). Siegler has suggested that this increased ability to solve problems may be due to more than just increased knowledge, but also to an increase in "...a more general learning ability" (p.117). For example, in the scale balance problem, 8-year old children were better able to benefit from experience with the problem than 5-year olds even though their initial knowledge of the problem was the same. This was due to better encoding skills of older children (Siegler, 1976). The purpose of this paper was to investigate the development of another skill that is part of mature problem solving, making inferences while trying to solve a novel problem.

Since a novel problem is one that the problem solver has not previously encountered, the first attempts to solve it may be wrong. Therefore, in order to develop a successful strategy, the person must make inferences about why the answer is wrong and about how to revise the strategy in order to solve the problem. The problem solver may make such inferences both during problem solving and after an answer, particularly an incorrect one, has been obtained. Most research studies investigate the development of inferences that are made after an answer is obtained. It is known, for example, that as children get older they get better at making inferences about how to solve a problem after experience with it (Siegler, 1976; Steinberg, 1980). The focus of the present study was the inferences children make while doing a problem, in contrast to after an answer has been obtained.

Before continuing, it is necessary to define the terminology to be used in this paper. Solving the problem here means developing a strategy that will lead to the correct answer for the entire set of such problems. Getting the correct answer to some problems and not to others does not mean solving the problem in the sense intended here. The expression "during problem solving" refers to the period when the student is in the process of working toward a solution or a single problem, before the final answer is achieved.

One of the reasons a person makes inferences during problem solving, before completing the problem, is to judge whether one's current state is a dead end, or likely to achieve the desired goal. For example, a person may be able to make an inference about whether there is sufficient information to solve the problem. If not, the problem solver can modify the strategy at this time and try again to solve the problem, rather than continuing on knowing that the current strategy is not likely to produce the correct answer. The purpose of this research was to investigate this facet of making inferences during problem solving, judging sufficiency of the information that has been obtained.

Children may be able to judge sufficiency of information but younger ones may not know how to use the information they derive to solve the

problem. Steinberg (1980) found that all second graders and most kindergartners solved a problem in which the strategy was to derive sufficient information in a single step. However, she did not determine the point at which this came about nor whether there were developmental differences. Did the children spontaneously judge the sufficiency of information during problem solving or only as a result of experience getting some wrong answers? On the one hand time data seemed to indicate that children in both grades judged during problem solving, because they took significantly longer to complete the problem when they had insufficient information than sufficient. However, all of the kindergartners and half of the second graders did not solve the problem immediately, but only after at least 12 trials. If they were judging sufficiency of information, they did not immediately make the inference about the implications of that knowledge for solving the problem.

The purpose of this study was to gain further information about the development of judgments of sufficiency of information obtained while attempting to solve a problem. As children mature, do they spontaneously make such judgements during problem solving? Do they make such judgements if they have a prompt to do so? Do they use the information derived to develop a strategy to solve the problem?

Experiment 1

Method

Subjects. All of the 36 kindergartners and 42 second graders in a low socio-economic public school participated. Four kindergartners were dropped because of extended absence. Therefore data are reported for 32 kindergartners. A classroom of PLATO terminals was in the school, and all children worked regularly at the terminals, doing lessons as a part of their school instruction.

Materials. The experimental medium was PLATO, the University of Illinois system for computer-assisted instruction. The task was basically the same as in Steinberg's earlier work (1980). In the task, two pictures were displayed simultaneously on the display screen. They were alike in all respects but one. For example, one picture showed a camel with its head down, the other with its head up (Figure 2, see p. 55). Simultaneously, a third, "covered" picture was shown on the screen and purported to be identical to one of the other two. The problem is solved in two steps: uncover half of the covered picture ("peek" response), and then decide which picture it will match for sure ("match" response). Obviously, the correct strategy is to uncover the discriminative part. The twelve pairs of pictures used in the earlier experiments were employed, and eighteen additional pairs of pictures were generated, for a total of thirty pairs. The pictures were divided into three, ten-item groups. In each group, half of the pictures were top-bottom different, and half were left-right different.

A ten-item discrimination pretest was generated. Each item consisted of three pictures, two alike and one different. The student had to identify the one that was not the same.

Design. Overall, there were two grades, two Option conditions (presence or absence of option to judge sufficiency of information) and two Guess conditions, (lucky guesses possible or not) for a 2 x 2 x 2 design. The reasons for these choices will be given below, following an explanation of the conditions.

In the Option conditions, the words "Can't tell" were displayed on the screen (Figure 6, see p. 59) after the student made the peek response. The experimenter read the screen for the children. For ease of communication, these words will be called the prompt. Thus students in Option conditions were allowed the option of indicating that they did not have enough information to tell for sure which picture the covered one would match. In the No option conditions, the students had to match one of the uncovered pictures even if they were aware that the information derived was not sufficient to make a decision.

In the Lucky guess conditions the student was able to make an incorrect peek response (uncover the nondiscriminative part) but make a correct match by luck. In the No lucky guess conditions, this was not possible. The computer chose the correct match after the student matched. If the student made an incorrect peek response, the match response was automatically wrong. For futher details about how this was possible, see the previous study.

The task structure in Steinberg's previous research could have misled students because, after a nondiscriminative peek, there was no opportunity for them to indicate that they had obtained insufficient information and could not make a match. This may have led them to believe that a correct match was possible but too difficult for them to do. If some of the children would have normally judged sufficiency of information during problem solving, but were so misled, then Option would enable them to solve the problem more quickly than No option.

It was also the case in the previous research that although second graders solved the problem in Lucky guess conditions, kindergartners did not. They succeeded only when no lucky guesses were allowed. Thus for second graders, Guess condition should not be a main effect. However, if kindergartners were really misled by the task structure as noted above, then Guess condition should not affect judging sufficiency during problem solving when prompt is available, but should make a difference when there is no prompt.

If the children can judge sufficiency of information, but do not know how to use this knowledge to solve the problem, those in Option conditions would use the option, but the number who are likely to solve the problem would be no greater than in No option conditions.

Procedure. Each child was randomly assigned to one of the 4 experimental conditions. The first session consisted of the discrimination pretest, an orientation to the task and then the 10-trial task. The second and third

sessions were held on consecutive school days and each consisted of a 10-trial task.

The experiment was conducted while the participants were doing regularly scheduled lessons on PLATO. The child was asked to leave his or her terminal and come to one in another part of the room, with no other students next to it, to do a "special game." The experimenter was seated next to the child. For each trial, two pictures appeared side by side on the screen and a "covered" one below them. The student decided which part to peek at, and touched it. The computer "uncovered" that part and displayed the appropriate part of the picture. At this point the procedure differed for students in Option and No option conditions.

Students in No option conditions then decided which picture the partially covered picture would match for sure, and touched it. They had to pick a view, even if they knew they could not tell. The computer then "uncovered" the rest of the covered picture.

For students in Option conditions, after peeking, the computer not only uncovered the chosen part, but also displayed the prompt, "Can't tell" in large letters at the bottom of the screen. Students could either match the prompt or one of the uncovered pictures above. If the student chose the prompt, nothing further happened, and the student went on to the next problem. If the student matched one of the pictures, the procedure was the same as in the No option conditions: the computer uncovered the rest of the picture.

In the Lucky guess conditions, the view that was displayed under the cover was chosen randomly before the student peeked, subject to the constraint that not more than three successive trials would match the picture on the same side of the display (left or right). Thus, the student could peek at the nondiscriminative part of the covered picture and still match the correct picture by luck. In No lucky guess conditions, if the student peeked at the nondiscriminative part, the computer made sure the student's match was wrong. If the student matched the left picture, the computer then generated the right hand one as the view under the cover, and vice versa.

In all conditions, neither the experimenter nor the computer judged the match to be right or wrong. It was the child who judged whether the match was correct or not. The entire task was self-paced in all three sessions.

Results

Mean score on the discrimination pretest for kindergartners was 8.1 out of 10, with standard deviation 1.5. Therefore, kindergartners' initial knowledge of the problem was judged to be adequate.

The number of correct peeks, rather than the number of correct peeks that were followed by a correct match, was chosen as the dependent variable, for two reasons. First, the choices for the peek response were

exactly the same for all experimental conditions. In every condition, the student first had to decide which part of the covered picture to peek at, and to touch that part. Once that was done, the student had to decide on a match response. At this point, as noted above, the choices were different and not strictly comparable. Second, it was demonstrated in the earlier study that if a student peeked correctly, the probability of his matching correctly, too, was .94. Hence the number of correct peeks was a very good measure of students' knowledge of how to solve the problem. Criterion of success in the task was a score of 80% correct peeks by the end of at least one, 10-trial session.

A significantly higher proportion of second graders than kindergartners reached criterion (Table 1), χ^2 (1) = 18.83, $p<.01$.

Table 1

Number of Successful and Unsuccessful Students

With and Without Option

Experimental Condition	Kindergarten		Second grade	
	Successful	Unsuccessful	Successful	Unsuccessful
No lucky guess				
With	2	5	9	2
Without	1	8	9	2
Lucky guess				
With	2	7	7	3
Without	4	3	8	2
Total	9	23	33	9

Separate analyses for each grade revealed that experimental conditions did not significantly affect the proportion of children who reached criterion.

Table 2 shows the mean number of correct peeks by grade and experimental condition. ANOVA showed a significant main effect of age $F(1,66) = 29.72$, $p < .01$. Second graders solved the problem more quickly than kindergartners. Neither experimental condition nor interactions were significant.

Table 2

Mean Number of Correct Peeks by Grade and Experimental Condition

Experimental Condition	Grade					
	Kindergarten			Second		
	n	Mean	S.D.	n	Mean	S.D.
No lucky guess						
With option	7	16.57	3.05	11	22.10	5.22
Without option	9	16.33	4.42	11	24.55	4.39
Lucky guess possible						
With option	9	16.67	2.29	10	22.10	5.55
Without option	7	18.00	4.04	10	21.90	5.11

Since the randomness of lucky guesses was not controlled, some of the students in Lucky guess were effectively in the No lucky guess condition during one or more of the sessions; that is, they were "unlucky" significantly more than half of the trials. When their peek was wrong, the match was wrong. In other cases, the total number of lucky guesses was not significantly different from .5, but the sequence was not random. So, for instance, the student might have three successive unlucky guesses in a row. This, too, may have biased the results. Therefore, an ANOVA was performed for only children in the No lucky guess conditions. Again, grade was a significant main effect, $F(1,34) = 21.78$, $p < .001$.

If a child did the first eight consecutive problems correctly, success was attributed to preplanning or already knowing the strategy. One of the nine successful kindergartners and six of the 33 successful second graders preplanned.

Considering all children who were in Option conditions, the mean proportion of times that kindergartners used the prompt when they should have (after a nondiscriminative peek), was .26 in No Lucky guess condition and .27 in Lucky guess. For second graders, the proportions were .24 and .34 respectively (Table 3a). Reanalysis by grade and success in solving the problem showed that the proportions of use of the prompt were .52 and .18 for successful and unsuccessful kindergartners, respectively. For second graders, the respective proportions were .31 and .19 (Table 3b). There was a significant main effect for success, $F(1,33) = 5.22$, $p < .05$. Neither

Table 3a

Mean Proportion of Appropriate Uses of Prompt,

by Grade and Experimental Condition

	Grade					
	Kindergarten			Second		
Condition	n	$\bar{x}$	S.D.	n	$\bar{x}$	S.D.
No lucky guess	7	.26	.184	11	.24	.204
Lucky guess	9	.27	.337	10	.34	.342

Table 3b

Mean Proportion of Appropriate Uses of Prompt,

by Grade and Success in Solving the Problem

	Grade					
	Kindergarten			Second		
Condition	n	$\bar{x}$	S.D.	n	$\bar{x}$	S.D.
Successful	4	.52	.360	16	.31	.305
Unsuccessful	12	.18	.189	5	.19	.131

grade nor interaction were significant in this analysis. Successful kindergartners and second graders, that is those who reached criterion, used the option correctly on a greater proportion of trials than unsuccessful students.

One kindergartner used the option correctly on every trial when she should have and also, incorrectly, on 6 trials when she had sufficient information. Since she ultimately reached criterion, her success was attributed to the availability of the option. Although she may not have completely comprehended the prompt initially, she obviously figured it out by the end of the last session.

All of the other three successful kindergartners used the option inconsistently. None of them used it after the first incorrect peek. Two of them did not even use it at all until the second 10-trial session.

Second graders' use of the option was different. They used the option after the first incorrect peek, and on early trials. Their low usage was on later trials.

Discussion

Second graders performed significantly better than kindergartners in both the number of correct peeks and the proportion of children who reached criterion. Children in both grades made little use of the option to judge sufficiency of information, but for different reasons.

Kindergartners did not judge the sufficiency of information during problem solving even when they had the option to do so. Overall, they used the option only 25% of the time that it was appropriate. Successful kindergartners used the option with 52% accuracy. However, their success cannot be attributed to the fact that the option enabled them to follow a spontaneous inclination to judge the sufficiency of the information they had obtained. Three of the 4 successful children in Option conditions used the prompt only after they found that some of their final, match answers were right and some were wrong. Furthermore, four children in No option conditions also solved the problem. With one exception, successful kindergartners modified their problem solving strategies after completing the problem, not while doing it.

Second graders also made little use of the option when they should have. However, the reasons are different than for kindergartners. Second graders understood the meaning of the option. Most of them used it correctly after the first incorrect peek. Then they did a long succession of correct peeks. However, if on later trials, they made an incorrect peek, they tended not to use the prompt. An explanation, supported by their teachers, is that they knew when they couldn't tell for sure, but wanted to see if they could guess right anyhow.

Unsuccessful students in both grades made no inference about the purpose or the usefulness of the words, "Can't tell". One explanation is

that they were unable to judge sufficiency of information. It is a skill they had not acquired. Another explanation is a "cultural" one. In elementary school, particularly in the primary grades, the students are expected to give an answer to each question. It is not often, if ever that children are asked to judge whether they have sufficient information to make a response.

One further developmental difference was the number of chidren who preplanned the strategy: only 1 of the 9 successful kindergartners, compared to 6 of 33 second graders.

Experiment 2

Purpose Kindergartners did not just make the inference, from "Can't tell," that they should judge their response during problem solving as well as after they get a final answer. Could they solve this problem if the information they obtained during problem solving was judged for them; that is, if the computer told them that the peeks were right or wrong? The purpose of Experiment 2 was to answer this question.

Method

Subjects. The children were 29 of the same 32 kindergartners who had participated in the first experiment, and who were still enrolled in the school.

Materials. The stimuli were changed in the experiment because the pictures in Experiment 1 may have misled the kindergartners to attend to irrelevant attributes of the problem. Some, for example, liked the engine of the train and peeked at that part even though it was not discriminative. Therefore, the pictures were replaced by 2-digit numbers, but the display format was the same. The numbers differed in either the tens or units place. Numbers were generated randomly by the computer, subject to the constraint that there was no duplication of digits. So, for example, a number like 22 was not allowed.

Design and procedure. There were 2 conditions, control and experimental. The control condition was like the No-option, no lucky guess condition of Experiment 1. In the experimental condition the computer judged the peek response: After the peek response, the screen displayed "GOOD" or "OOPS" below the peek, depending on whether the peek gave sufficient information (right) or not (wrong). After the match response, the computer gave feedback for both conditions. Students saw a "very good" or an "oops" on the screen.

Half of the children from each of the experimental conditions in Experiment 1 were randomly assigned to the current control condition; the other half to the experimental condition. The student was not restricted to a fixed number of trials in each session, as in Experiment 1, but rather worked to criterion or until they became tired or bored, whichever came first. Children who did not reach criterion during the first session, were given the opportunity to try again the next school day. This change was

made because in the first experiment some of the children seemed to be getting the idea just when the 10 trials in the session were completed. It seemed that they might have done better if they were allowed more trials in a session.

Criterion was decreased to 5 successive correct peeks. When the child reached criterion, a ribbon and congratulatory message were displayed on the screen.

Results

The dependent variable was the trial of last error (TLE), that is the last trial they did wrong before they did 5 problems in succession correctly. Ten of the 13 children in the experimental condition and 12 of the 16 controls solved the problem. Mean trial of last error for successful students was 20.7 for experimental and 17.667 for control conditions. The difference was not statistically significant.

Twenty-two out of 29 children solved the problem compared to 9 out of 32 in the first experiment. The difference was significant, χ^2 (1) = 12.03, p = .0005.

Discussion

Most of the kindergartners solved the problem. However, they performed equally well whether or not the computer judged their peek response for them. They did not make the inference from this feedback that they should evaluate and, where appropriate modify the strategy at this point in the solution process. The fact that a significantly greater number of kindergartners solved the problem in this experiment than in the preceding one can be attributed to one or more of the following variables: student maturation, stimuli free of irrelevent attributes, individualizing the number of trials allowed in each session and to lower criterion.

The summarize, kindergartners did not judge sufficiency of information during problem solving. Those who succeeded did so only after they completed a succession of problems and discovered that some of the answers were right and others were wrong. Their goal was to get the answer. They did not consider the possibility that their strategy might need modification until after it was implemented and resulted in some wrong match answers. In contrast, second graders judged sufficiency of information during problem solving. They were able to forgo completing the problem before revising the strategy and trying to solve the problem again.

Feedback Compared to Practice Alone as Instruction in Problem Solving

Esther R. Steinberg

University of Illinois
at Urbana-Champaign

The author gratefully acknowledges the cooperation of the principals, teachers, and children in the Champaign public schools who participated in this research.

Abstract

Ninety-three first and third graders children tried to solve a novel reasoning problem on PLATO, in one of three conditions: experience only, feedback that was a visual record of the student's work on the first 5 problems, or interactive instructional feedback on how to solve each problem that the child had done wrong. Results showed that experience alone was as effective as the feedback conditions in terms of both the number of successful students and the number of trials needed to succeed. Reasons are discussed for failure of feedback conditions to be more facilitative than experience alone. Successful students in all conditions generalized their strategies to two transfer tasks, but transfer was not immediate.

Feedback Compared to Practice Alone as Instruction in Problem Solving

The purpose of this research was to compare three conditions, each using some unique feature of computer-assisted instruction, for effectiveness in helping children solve nonroutine problems. The conditions were: experience only, feedback that was a visual record of the student's work on the first 5 problems, and interactive instructional feedback about how to solve a problem that was done wrong. In the Experience condition, the computer was programmed so that students could not get the right answer if the problem solving strategy was wrong. This structure was maintained in both feedback conditions. In one feedback condition, Record, the computer also displayed on the screen the first 5 problems the student had done, along with the student's responses and whether they were right or wrong. In the other condition, Instruction, the computer interactively helped the student do the problem correctly if it had been done wrong.

The focus here was on problems for which the students had to generate a solution strategy. They were not problems that could be solved by applying a set of rules. When confronted with such problems, inexperienced problem solvers and beginners in a subject area either do not explicitly formulate a solution strategy (Mosher & Hornsby, 1966) or they only do a limited amount of planning (Reif, 1977; Chase & Simon, 1975). Therefore, their first attempts to solve the problem frequently result in wrong answers. Although some students may be able to benefit from experience alone and generate a solution strategy, it is not generally the case that "practice makes perfect." If it were, then there would be no need for further instruction. The question then arises, what kind of feedback should be provided to help students generate a successful strategy?

The two feedback conditions that were developed here were an extrapolation of research about feedback in written instruction. In a review of that research, Kulhavy (1977) noted that feedback is helpful when it confirms correct responses, that is, lets the student know how well he is doing, and either enables the student to identify and correct the errors or corrects the errors for the student. In problem solving, a technique that is helpful in identifying errors is making inferences from the results of previous attempts to do the problem. Immature problem solvers may not remember what they did previously, or may not realize that this is a useful strategy. CAI is particularly suited to keeping a record of the student's previous work. Thus one form of feedback that might be facilitative would be to show the student the problems he did previously, how he did them, and whether they were right or wrong.

In problem solving, feedback that is parallel to correcting the student's errors in written instruction is leading the student through the correct procedure in individual problems that the student did wrong. PLATO is also well suited to provide this type of individualized interactive feedback.

Some students can develop successful strategy solutions from experience alone, particularly if the task is structured to maximize this

possibility. This was the case in my previous work (Steinberg, 1980) and will be explained in detail below. Therefore, the Experience condition was included in this research to determine whether practice accompanied by feedback was really more effective than practice alone.

The task chosen for the present research was the same as that in my previous work. The structure of the task employed here evolved from that research. Therefore it will be reviewed here. The problem itself was a modification of a PLATO reading lesson (Obertino, et al., 1977). The objective was for the student to figure out which of 2 pictures was hiding under a cover, by obtaining a minimum of information. Two views of a common object were displayed on the screen. They were alike in one part and different in the other. For example, two dragons were shown, with exactly the same body, but one was breathing fire and one was not. Simultaneously, a third, covered picture was shown on the screen, and purported to be identical to one of the uncovered pictures. The student's task was to uncover just one half of the covered picture and on that basis decide which of the pictures it would match. The strategy, of course, was to uncover the discriminative part (the head in the example above).

This was a difficult task for kindergartners. In the initial experiment, the view that was under the cover was randomly picked by the computer before the student began the problem. Therefore, the student could uncover the nondiscriminative part and still match the correct picture by luck. Under this condition, kindergartners were unable to solve the problem. That is, although they did some problems correctly, they were unable to generate a strategy for doing all such problems correctly. The task was restructured so that the computer did not choose the view that was covered until after the student decided which part to uncover. If the student chose to uncover the nondiscriminative part, the computer was programmed to make the student's match response be wrong. If the student matched the picture on the left, the picture under the cover turned out to be like the one on the right, and vice versa. Under this task structure, experience alone was enough for two-thirds of the kindergartners to generate a successful strategy. The question then was, what kinds of additional feedback would help the unsuccessful students solve the problem.

The purpose of the experiment then, was to compare the effectiveness of these three conditions, Experience, Record of previous work, and direct Instruction for the specific problem. How did they compare in terms of the total number of students who could solve the problem? the number of trials needed to generate a strategy? the generalizability of what the student learned?

Experiment

Method

Subjects. All of the 57 first graders, 30 girls and 27 boys, in a public school in a middle class socio-economic income area participated. Their ages ranged from 70 to 89 months, with mean 77.6 months. In addition, 36 third graders, 18 girls and 18 boys from the same school were

the 4 third grade teachers to select 9 students for the experiment, 3 each who were below, at grade level, or above in arithmetic.

Materials. The research medium was PLATO, the University of Illinois system for computer-assisted instruction.

Three categories of lessons were employed: (1) a discrimination pretest, (2) fun activities and arithmetic lessons that were available on the system, and (3) the tasks for the experiment.

The discrimination pretest was one that had been used in the previous research, and was employed to be sure that the children did have the prerequisite discrimination skills to do the problem. In addition it served to review the concepts of same and different, which were used in the problem. The fun lessons were used to acquaint the student with PLATO, and as a reward for doing the problems in the experiment.

The basic task was that as described above, but the stimuli in the initial task were numbers rather than pictures of familiar objects. In previous research (Study 2) kindergartners were more successful with numbers than with the pictures.

In addition, there were 2 transfer tasks. One used different stimuli, the 30-item set of pictures from previous experiments. The second transfer task differed in the number of digits and the number of choices. The stimuli were 3, 3-digit numbers. In all of the tasks that involved digits, there were no double or triple-digit numbers, such as 44 or 777. The three-digit numbers were generated so that all were alike in one position, two were alike in a second position, and they were all different in a third position. An example is the set, 345, 748, and 741.

Design. Children were randomly assigned to one of three conditions: Experience, Record, or Instruction. The order of the tasks was the 2-digit, followed by the pictures, and then the 3-digit tasks. The children did each task until criterion, 8 successive correct trials (problems), or until the maximum number of trials was completed, whichever came first.

In order to obtain a baseline of initial skills, all students did the first 15 problems in the initial, 2-digit, task in the Experience condition. Students who reached criterion before the 15 problems were completed, went on to the transfer tasks. Those who did not reach criterion in the first 15 trials then continued on in the 2-digit task, in their assigned experimental condition, until they reached criterion or completed a total of 99 trials, whichever came first. They then did each of the transfer tasks in Experience condition, working to criterion or until the maximum number of trials was exhausted (30 pictures and 60, 3-digit numbers), whichever came first. All transfer tasks were done in Experience condition.

Procedure. The PLATO terminal was in an empty classroom in the school. Each child came to the classroom to participate in the experiment. During the first session the student was acquainted with PLATO. The child

did some PLATO fun lessons, such as making funny faces. First graders then did the discrimination pretest. Third graders did not do the discrimination pretest, since it was assumed that by this grade they had the skill. They continued on as in the first graders' second session.

At the beginning of the first graders' second session, the computer randomly assigned the child to an experimental condition. Then the child did a lesson that was an orientation to the task, with the 2-digit numbers as stimuli. This was followed by the task itself, in Experience condition for all students. and a maximum of 17 trials. The student first did two "warmup" trials, and data for them were not included. The student then worked to criterion or until the 15 additional trials were completed, whichever came first. Criterion was defined as 8 problems in a row correct.

At the next session, children who had not reached criterion continued doing the initial task in their assigned experimental condition until criterion or until they became restless, whichever came first. In all conditions, PLATO allowed the students to get the match answer right only if the strategy (i.e. the peek response) was right. In Experience condition, the computer did not judge either the peek or the match answer. The student had to judge for himself that it was right or wrong.

In Record condition, PLATO also gave feedback after the peek response. It drew a star and wrote, "Good" if the answer was correct, and wrote, "No" if it was wrong. In addition, PLATO displayed each problem at the top of the screen, along with the student's peek response and OK or NO (Figure 7, see p. 60). There was space for a record of only the first five problems. The experimenter explained the meaning of the record to the student and urged the child to try to figure out why some of them were NO and some OK.

In Instruction condition, PLATO gave the same feedback on peek responses as in Record condition. In addition, after the student did a problem wrong, the screen was erased and the problem displayed again. Then the experimenter pointed to the part of the covered number that was the same digit in both of the uncovered numbers. For example, if the numbers were 28 and 38, the experimenter pointed to the right part of the covered number and said, "What number is under there?" The student responded. In some instances the child answered wrong. For instance, for the numbers 28 and 38,the child might say that the digit under the right part of the covered number could be a 3, or else a 3 or an 8. Then the experimenter pointed to right hand parts of 28 and of 38 and asked the student, "What number is in this part?" After the student answered, the experimenter explained that therefore 8 was the only digit that could possibly be under the cover on that side. Then PLATO wrote that number (here 8) and the words, "Not here," below the covered part and drew a finger pointing to the other part and wrote, "Touch here" (Figure 8, see p. 61). The experimenter read the message, since not all first graders could read it at the time of testing. The student then peeked at the correct part and was able to match correctly without assistance.

At succeeding sessions, each child worked until criterion, until 99 trials were completed, or until the child became bored, whichever came first. After completing the 2-digit task, the child did the pictures task until criterion or until all 30 picture sets were exhausted. Finally the child did the 3-digit numbers task, again until criterion or a maximum of 60 trials.

If a child reached criterion on a task, PLATO displayed a smiling face and the message, "VERY GOOD, (student's name). You are all done. GOODBYE."

At each session the child did the experimental task first and then chose fun lessons on PLATO.

Results

Mean score in the discrimination pretest was 9.2 (S.D.=1.139). Only 2 children scored below 70% and both of them subsequently solved the problem.

Mean number of sessions per student were 8.9, 9.6, and 9.1 for Experience, Record, and Instruction Conditions, respectively. The respective means of the total number of minutes per pupil were 56.1, 61.0, and 62.2. ANOVA revealed no significant differences between conditions in either number of sessions or total time to completion.

Mean scores for the first 15 trials were computed. Some children reached criterion in fewer than 15 trials, and consequently their scores were "artifically" low. Since it was assumed that children who reached criterion before completing all 15 could also do the rest of the problems correctly they were so credited. For example, if the child reached criterion after 12 problems, the score on the first 15 was computed as the number of those 12 that were correct plus the remaining 3 of the 15 that the child did not have to do. Mean scores for first 15 problems were 9.15, 8.95, and 8.316 for first graders in Experience, Record, and Instruction conditions, respectively. The differences were not significant. The groups were therefore considered comparable in initial ability.

The numbers of students who succeeded in each condition, by grade and task are shown in Table 1. For 2-digit problems, there were no significant differences between conditions for either first or third graders. Collapsing across all conditions, and comparing grades, a significantly greater proportion of third graders than first graders succeeded, $X^2(1) = 12.2$, $p = .0005$.

For each task chi squared was employed to compare the proportions of first graders succeeding in each condition. None were statistically significant. Simliarly, within each condition the proportions succeeding in each task were compared. None of these were significantly different.

As expected, with just 1 exception, all third graders succeeded. Therefore there were no condition effects.

Table 1

Number of Students Who Did and Did Not Reach Criterion in Each Experimental Condition by Grade and Task.

Condition	Task					
	2-Digit		Pictures		3-Digit	
	Did	Did Not	Did	Did Not	Did	Did Not
	First grade					
Experience	14	5	10	9	15	4
Record	10	9	12	7	14	5
Instruction	14	5	9	10	14	5
	Third grade					
Experience	11	1	11	1	11	1
Record	12	0	11	1	12	0
Instruction	12	0	11	1	11	1

Mean trial of last error (TLE) by grade, experimental condition, and task are shown in Table 2. Grade and task were both significant main effects (Table 3), as was the interaction of grade by task. Third graders solved the task in fewer trials than first graders. Experimental condition was not significant. None of the other interactions were significant.

Since there were significant differences between the grades, ANOVA was calculated separately for each one. There were no significant main or interaction effects for third graders. Although they solved the transfer tasks in fewer trials than initial task, the difference was not statistically significant.

ANOVA of TLE's of first graders showed a significant main effect of task (Table 4); TLE was higher for 2-digit than for transfer tasks. The interaction beween condition and task was also significant.

Table 2

Trial of Last Error for Children Who Succeeded on Each Task by Experimental Condition and Grade

	Task								
	2-Digit			Pictures			3-Digit		
	First Grade								
Condition	n	$\bar{x}$	S.D.	n	$\bar{x}$	S.D.	n	$\bar{x}$	S.D.
Experience	14	25.1	28.257	10	10.1	8.925	15	12.8	12.723
Record	10	11.2	16.605	12	9.5	8.868	14	12.4	16.008
Instruction	14	36.1	26.558	9	6.9	9.239	14	8.3	9.101
	Third Grade								
Experience	11	7.3	13.267	11	6.4	8.237	11	2.9	4.989
Record	12	8.3	10.974	11	3.6	6.918	12	3.6	4.420
Instruction	12	10.4	13.235	11	4.0	6.261	11	6.4	11.012

Data were collapsed across conditions to compute mean TLE for each task. They were 25.5, 8.936, and 11.210 for Experience, Record, and Instruction conditions, respectively. Duncan new multiple range test (Kirk, 1968) showed that the mean for the 2-digit task was significantly higher than for pictures, $W_3 = 10.600$, $p < .01$,) and higher than for the 3-digit tasks ($W_3 = 10.804$, $p < .01$). The difference between the pictures and 3-digit tasks was not significant.

It was considered possible that the lower TLE in the pictures task was not due to learning but rather a reflection of the fact that fewer trials were allowed in the pictures problem than in the 2-digit, (30 compared to 99, respectively). Some students might have succeeded if they were allowed more trials, in which case the mean TLE would have been higher. However, most of the children who failed the pictures task had previously failed the 2-digit task (3 of 5 in Experience condition; 7 of 9 and 5 of 5 in Record and Instruction conditions, respectively). There was no reason to believe that they would have succeeded had they had more trials. To summarize, TLE's were significantly lower in both transfer tasks than in initial task. However, transfer was not immediate to either one.

Table 3

ANOVA, Trial of Last Error by Grade, Task, and Experimental Condition

Source	MS	df	F	p
Grade (G)	4102.0131	1	20.46	<.0005**
Task (T)	1978.7059	2	9.87	<.0005**
Condition (C)	276.6761	2	1.38	.254
G x T	638.9136	2	3.04	.050*
G x C	117.0125	2	.58	.559
T x C	416.4437	4	2.08	.085
G x T x C	394.5065	4	1.97	.101
Within Cell	200.4958	196		

Table 4

ANOVA, First Graders' Trial of Last Error by Task and Experimental Condition

Source	df	MS	F	p
Condition (C)	2	377.6920	1.24	.293
Task (T)	2	2456.4108	8.07	.001**
C x T	4	804.1572	2.64	.038*
Within Cell	103	304.2734		

The significant interaction between task and condition was due to the fact that there was a decrease in TLE's for students in Experience and Instruction conditions, but not for students in Record condition. In the 2-digit task, students in Record condition had a considerably lower TLE than students in the other conditions. Data were analyzed further to determine whether this difference could be attributed to experimental

condition. The initial skills of students in each of the experimental conditions were reexamined. The groups were previously considered to be equivalent because the number of children who reached criterion was not significantly different. In addition, mean scores on the first 15 problems were not significantly different. However a different proportion of students in each condition succeeded within that 15-problem preliminary. That is, 7 of the 14 children in Experience condition who ultimately solved the 2-digit problem did so before trial 15. Comparable numbers were, 7 of the 10 successful students in Record Condition and 3 of 14 in Instructional condition (Table 1). The difference between the proportions just missed statistical significance, $\chi^2(2) = 5.81$, $p = .055$. This indicates that different proportions of students succeeded before actually getting the feedback. Therefore, mean TLE's were recalculated for children who succeeded after the first 15 trials, that is after getting the feedback of their experimental conditions (Table 5). The differences were not statistically significant. Apparently, the low TLE of initial task of children in Record condition can be attributed to the fact that a disproportionate number of them solved the problem before receiving the feedback.

Table 5

Mean Trial of Last Error for First Graders Who Solved 2-Digit Problem After Trial 15

Condition	n	$\bar{x}$	S.D.
Experience	7	30.00	27.809
Record	3	17.33	16.042
Instruction	11	29.636	23.235

The number of children who solved 1 or more of the tasks is given in Table 6. With 1 exception in 2 of the conditions, each of the children who solved the 2-digit also solved the 3-digit task. However, only 28 of the 38 who were successful on 2-digit were also successful on the picture task.

Children's spontaneous comments were recorded along with responses to occasional questions by the experimenter. First graders did not differentiate between a problem being hard to do or impossible to solve. Twelve children spontaneously made comments such as, "Now it gets hard." or "This is tricky" when in fact it was impossible to solve the problem because their peek resulted in nondiscriminative information. There was a

Table 6

Number of First Graders Who Solved One or More Tasks

Condition	Problem-type Solved			
	2-Digit	2-Digit and Pictures	2-Digit and 3-Digit	All 3 Tasks
Experience	14	9	13	9
Record	10	10	10	10
Instruction	14	9	13	9

difference between first graders' and third graders' interpretation of the meaning of PLATO's comments "OK" and "NO". Fourteen first graders gave answers such as, "I don't know," or "I got it wrong." They did not analyze the reason for the error. In a similar way, children in Record condition were asked why PLATO had judged certain problems OK. (The problems and the peek responses were in full view.) Some gave responses such as, "That means I did it right." A few explained why. For example, in the problem that involved 71 and 91 one child said, "They both got 1's at the end." Third graders, on the other hand, analyzed the reason for the judgment. An "OK" means that the peek response enabled them to match the correct picture. After an incorrect peek, one child spontaneously commented, "Uh uh. There's two 1's in each number,"; another's comment, "I keep guessing one number when I should have guessed the other number." Third graders were more aware of what they were doing: "That's the third one in a row that's on that (i.e. left) side, but I better use logic."

Discussion

Experience alone, that forced the student's answer to be wrong when the strategy was wrong, was just as effective by itself as feedback which added a record of previous performance, or instructed the student on how to solve a problem that was done wrong. Memory deficiency was not the reason students failed to solve the problem. In order to benefit from Record feedback, the student had to infer the critical differences between problems done right and those done wrong. Among other things, they had to figure out how the wrong ones were different from the right ones, or at least what the wrong ones had in common. Since first graders did not analyze why an answer was wrong, beyond the fact that they did it wrong, such a record provided the student with no more information than a "NO."

What was surprising was the failure of the Instruction condition to be better than Experience. The student was told exactly what to do, and then asked to do problems that were almost identical. Even if the student was just following the directions blindly, he or she should have been able to

do the succeeding problems correctly. Some of the students had their own way of doing the problem, and although they might do one or two problems correctly after the feedback, they would revert to their old system, such as peeking at the part that was the same, or peeking in a preferred position like the right side. Their was a conflict between their natural strategy for doing the problem and the one they were taught by PLATO. Initially, they were unable to sustain that correct procedure. It took them just as many trials to reach criterion as the students who figured it out for themselves.

Just telling students what to do did not necessarily mean that the students would accept the instructed method over their own. They apparently did not understand why they were told to do the problem in the given way. PLATO did not explain that. It was left up to the student to infer the underlying concept, why the steps in the process led to the solution. First graders were unable to make that inference.

Feedback in problem solving helped students only to the extent that they were able to make the inferences from it that led to developing a solution strategy. To be effective in problem solving, feedback must do more than merely help the student get the right answers. Feedback should also help them generate a strategy. To accomplish that, the feedback should enable students to see the relationship between what they are taught to do and the goal they are seeking. A post hoc study, reported below, lends further support to this contention.

There were no significant differences between the conditions in the transferability of the strategies that the students generated.

Successful first graders solved both the picture and the 3-digit tasks significantly faster than they did the initial, 2-digit task. However, they did not solve either of these tasks immediately. One would have expected the student to require one or two problems as warmup for the pictures task, but certainly no more. The fact that the mean TLE was as high as 8.9 was indicative of either some difficulty with some of the discriminations or with an occasional abandonment of the logic of problem solving for an irrelevant attribute of the task such as a personal perference for a part of a picture.

If the students were generating a general strategy for the 2-digit problem, their TLE for the 3-digit task should have been no more than 1 or 2. Since the grand mean was 11.2, students obviously were unable to transfer their strategy directly. They obviously had to do some sort of strategy modification to solve the 3-digit problem. Interviews with the students revealed that in the 2-digit problems, they devised a strategy which was correct, but useful only for a 2-digit situation. The strategy was as follows. First decide which part is different. Then guess (to yourself) which digit is the correct one and peek at the appropriate part to see if you're right. Example: given 23 and 27, decide that the right part is different; guess it is the 7; peek at the right part of the covered number. If it is a 7, you're right and match the 27. If it is the 3, you can still match correctly even though your internal guess was wrong. In

the case of the 3-digit problems, this hypothesis testing breaks down. For example given the set, 345, 748, and 741, suppose the student tests the hypothesis that the first digit is a 3, and peeks at the left-most digit. Then if it is a 3, the student indeed knows that the covered number is 345. However, if the first digit turns out to be a 7, the student does not know if the covered number is 748 or 741. Thus the form of hypothesis testing that was suitable for the 2-digit task was not transferable to the 3-digit.

Revised Feedback and a Post Hoc Study

The main conclusion drawn from the experiment was that for immature students, problem solving feedback should relate the means to the goals and not merely tell the student what to do. Therefore, it was decided to develop this kind of feedback for the present task, and to test its effectiveness with students who had not previously solved the problem.

Method

In the experiment, some of the students had difficulty predicting just what numbers they could possibly see under each part of the cover. A few of them thought that if they peeked at the left side, they would see one of the digits of the number that was on the left. So, if the numbers were 23 and 27, a peek at the left side would reveal a 2 or a 3, and a peek at the right side would reveal a 2 or a 7. Preliminary instruction taught students how to predict the digits that could possibly be under each part of the covered number. The 3-digit task was used for training so that the student would have to think about a general strategy and could not get by with one that was specific to the 2-digit problem.

Feedback was revised to draw attention to the relationship between their strategy and their answers. This was accomplished in the following way. First the students had to state what numbers they could possibly see under each part of the covered number. This information remained on the display (Figure 3, see p. 56) while the student decided which part to peek at, and which number to match. If the peek, and consequently the match, were wrong the student did the same problem again, with the alternatives remaining in full view.

Nine first graders who had participated in the experiment, and had failed, were available for this study. Each child first practiced predicting what numbers could be under each part of the cover, for the 3-digit task. Then the child did the entire problem, with the possible numbers in each part displayed in that part. If the student did the problem wrong, PLATO hid a different one of the same, 3, 3-digit numbers and the student tried again. Thus the student had in front of him at all times the alternatives for each peek as well as the evidence that the match answer was right or wrong. The child did this 3-digit task until he could do 2 successive problems correctly on the first try.

Results

After completing 2 successive trials correctly, the child began the 2-digit task as in the experiment, in Experience condition. Seven of the 9

children solved it immediately, with mean TLE 2.4. Six of the seven also solved the pictures task. Five of the seven solved the 3-digit task.

Even the most immature children were able to solve the problem, not because they were told what to do, but because the feedback enabled them to infer the relationship between what they were doing and the results they were getting.

Discussion

It would have been faster to explain the relationship to the students than to develop this instruction. However, such an explanation may or may not have been as effective as the type of feedback that was provided. In fact, a question that needs futher research is whether you can explain the relationship between means and goals to students. It may be that students must figure that out for themselves and the instructor's role is to arrange the conditions that make this possible.

The subjects in these studies were children, but the conclusions may very well apply to adults who are novices in problem solving.

References

Brainerd, C.J. Cognitive development and instructional theory. Contemporary Educational Psychology, 1978, 3, 37-50.

Brown, A. L. & DeLoache, J. S. Skills, plans, and self-regulation. In R. S. Siegler (Ed.), Children's thinking: What develops? Hillsdale, NJ.: Lawrence Erlbaum, 1978.

Bryant, P. Perception and understanding in young children. London: Methuen, 1974.

Case, R. A developmentally based theory and technology of instruction. Review of Educational Research, 1978, 48, 439-463.

Chase, W. G. & Simon, H. A. Perception in chess. Cognitive Psychology 1973, 4, 55-81.

Gagne, R. M. & White, R. T. Memory structures and learning outcomes. Review of Educational Research, 1978, 48, 187-222.

Gelman, R. Cognitive development. Annual Review of Psychology, 1978, 29, 297-332.

Gholson, B. & McConville, K. Effects of stimulus differentiation training upon hypothesis strategies and stereotypes in discrimination learning among kindergarten children. Journal of Experimental Child Psychology, 1974, 18, 81-97.

Inhelder, B. & Piaget, J. The early growth of logic in the child. New York: W. W. Norton, 1964.

Kirk, R. E. Experimental Design: Procedures for the behavioral Sciences. Belmont, CA.: Brooks/Cole, 1968.

Klahr, D. Goal formation, planning, and learning by pre-school problem solvers or: "My socks are in the dryer." In R. S. Siegler (Ed.), Children's thinking: What develops? Hilldale, NJ: Lawrence Erlbaum, 1978.

Kulhavy, R. W. Feedback in written instruction. Review of Educational Research, 1977, 47, 211-232.

Mosher, F. A. & Hornsby, J. R. On asking questions. In J. S. Bruner, R. R. Olver, P. M. Greenfield, et al., Studies in cognitive growth. New York: Wiley, 1966.

Obertinc, P., Fillmen, L., Gilfillan, J., Silver, D., & Yeager, R. Elementary reading on PLATO IV, 1977. Urbana, Ill.: University of Illinois, Computer-based Education Research Laboratory, 1977.

Paiget, J. Judgment and reasoning in the child. Totowa, NJ.: Littlefield, Adams & Co., 1976.

Resnick, L. Task analysis in instructional design. In D. Klahr (Ed.), Cognition and instruction. Hillsdale, NJ.: Lawrence Erlbaum, 1976.

Siegler, R. S. Three aspects of cognitive development. Cognitive Psychology, 1976, 8, 481-520.

Siegler, R. S. The orgins of scientific reasoning. In R. S. Siegler (Ed.), Children's thinking: What develops? Hilldale, NJ.: Lawrence Erlbaum, 1978.

Steinberg, E. R. Evaluation processes in young children's problem solving. In press, Contemporary Educational Psychology 1980.

One of the camels is hiding under the cover. ●4
Touch the part you want to peek at.

Figure 1. Sample display of pictures task.

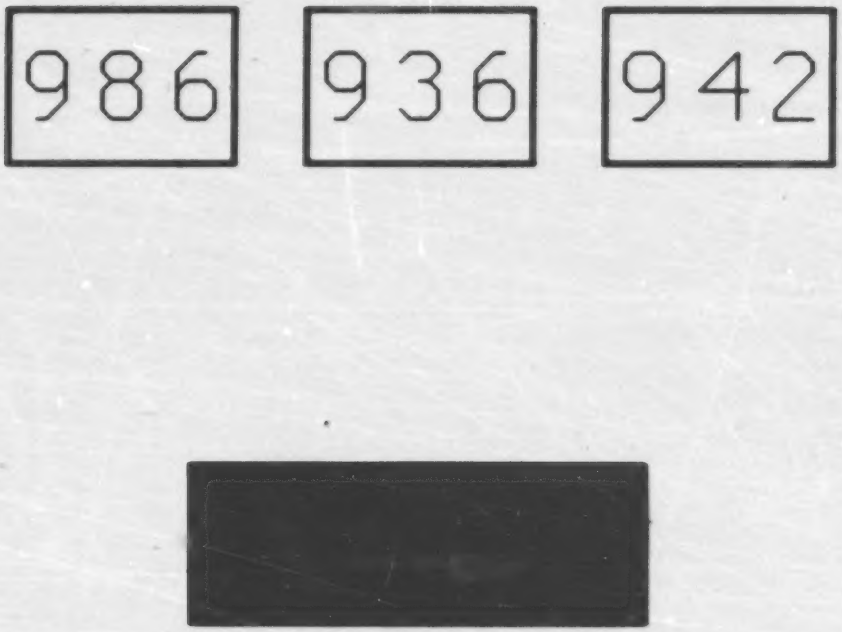
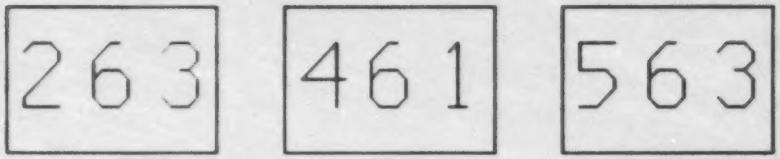


Figure 2. Sample display of three-digit task.

One of these numbers is hiding under the cover.



I hid a new number.

What box do you want to peek at? Touch it.

Figure 3. Sample display in post hoc study.

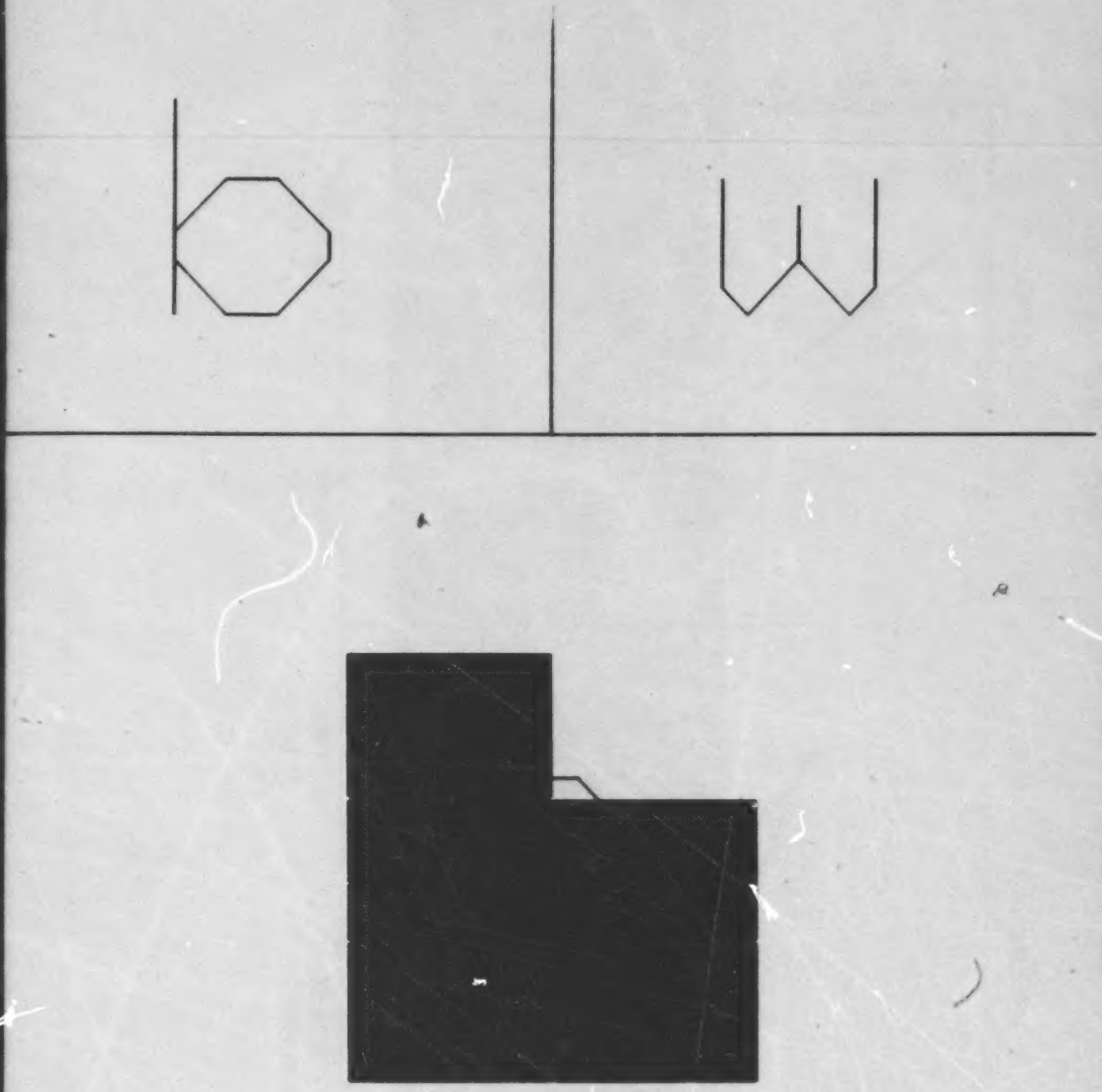


Figure 4. Sample display of reading task.

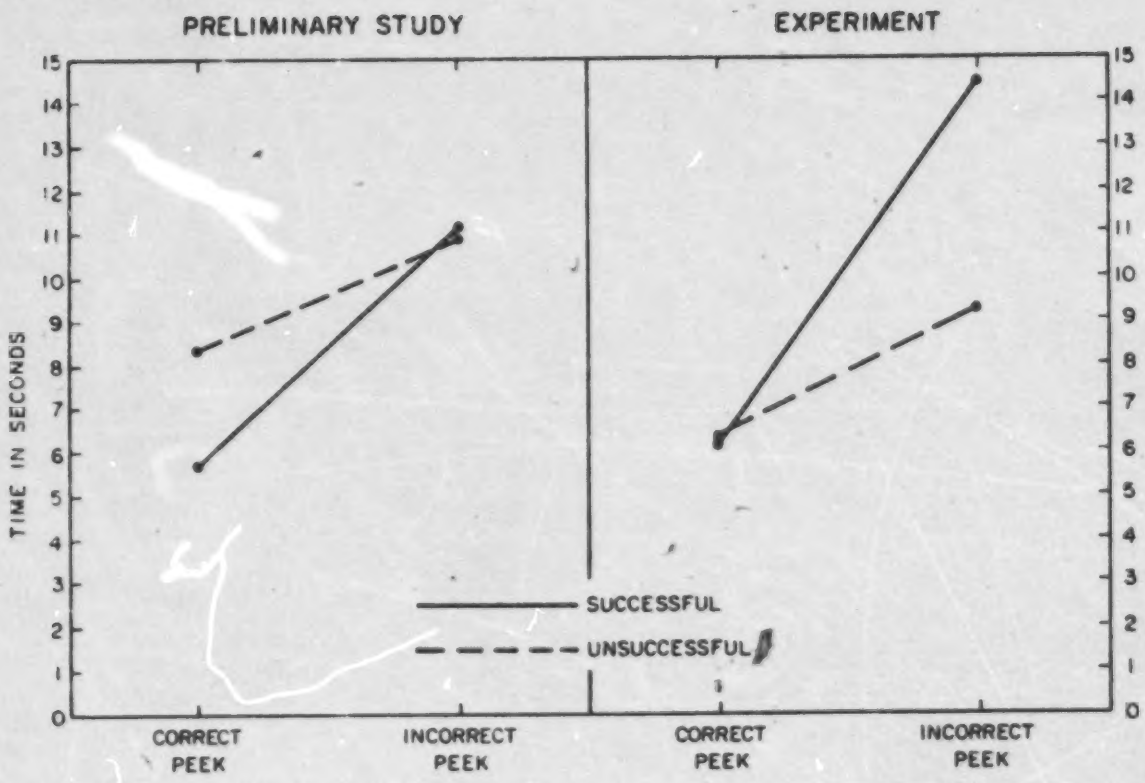


Figure 5. Latencies for successful and unsuccessful students.

Item 3

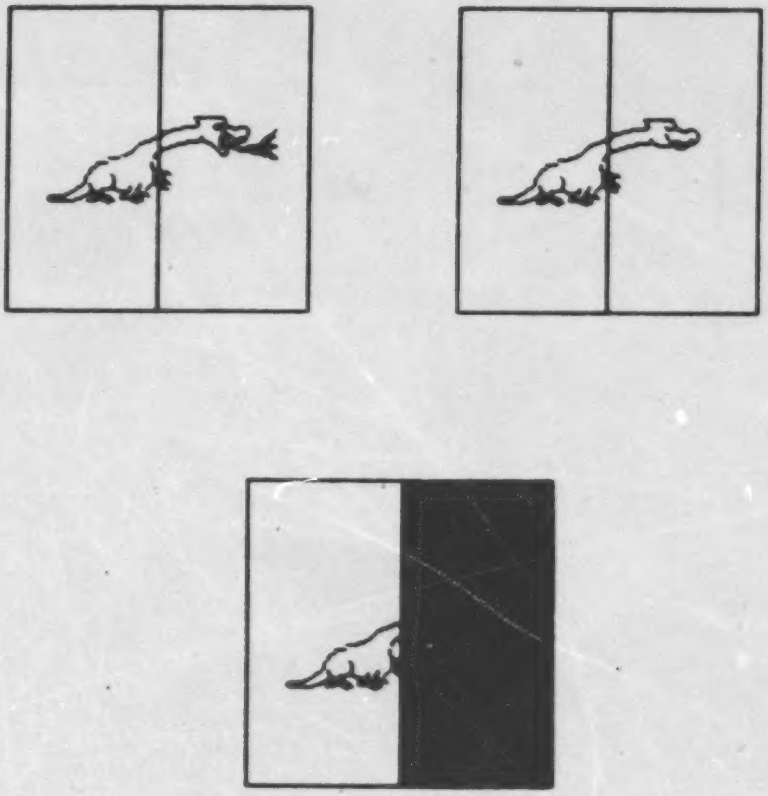
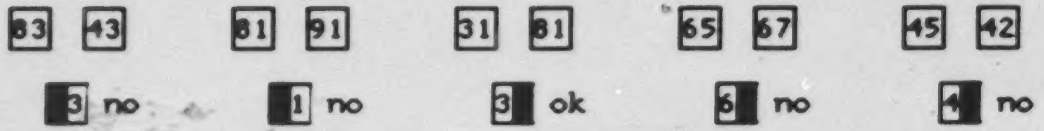
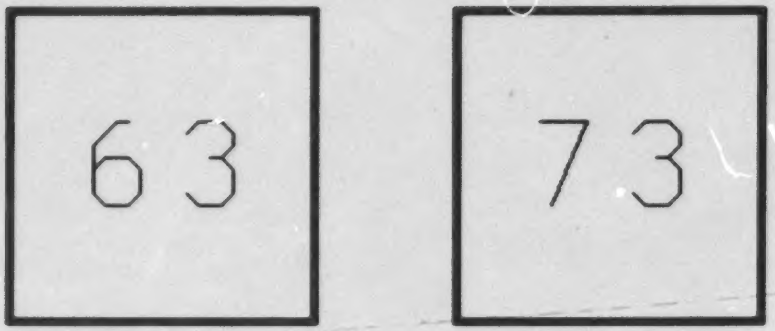


Figure 6. Sample display in Option condition.





A number is hiding under the cover.
Which part do you want to peek at? TOUCH

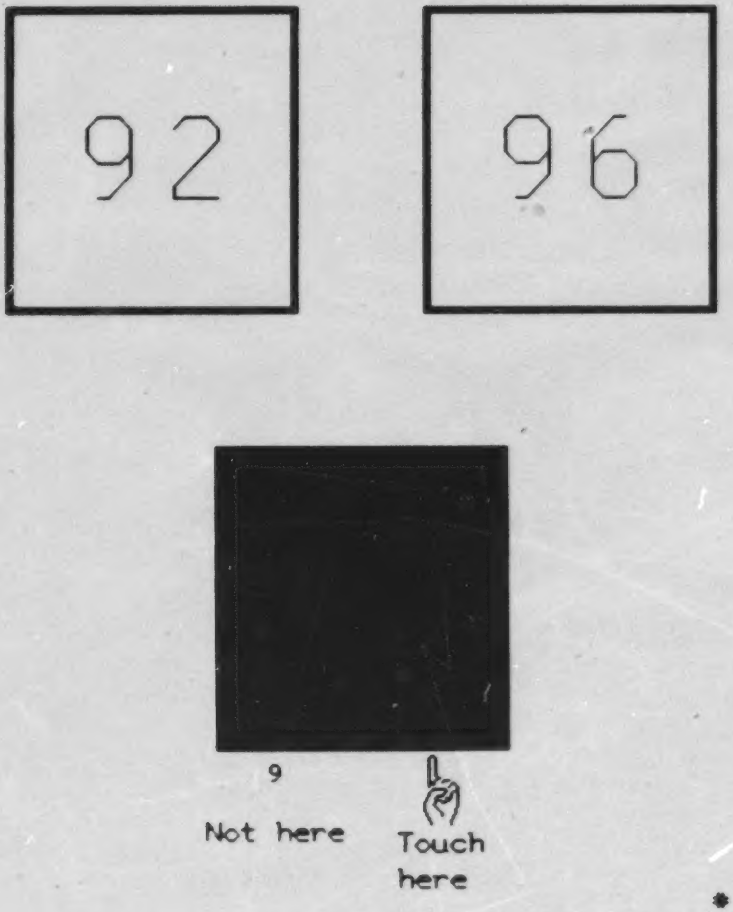


Figure 8. Sample display of Instruction condition.